Unconscious influences of attitudes and challenges to self-control

Deborah L. Hall

Duke University

B. Keith Payne

University of North Carolina, Chapel Hill

Please send correspondence to:
Keith Payne
University of North Carolina, Chapel Hill
Department of Psychology
Campus Box 3270
Chapel Hill, NC 27599
Email: payne@unc.edu

Remember the last time you were stuck in traffic, only to find after an eternity of inching, there was nothing but a fender bender neatly off to the side of the road to explain the delay? It's maddening to realize that the whole delay resulted from one driver after the other slowing down to take a good look. This realization usually occurs precisely as you are craning your neck, slowing down to take a good look. It's hard not to be captivated by the details of people's misfortune—just ask the folks who read gossip columns, watch horror films, or study social psychology.

Social psychologists have always been excited by findings that are nothing but bad news for our cherished ideas about how humans operate. Classic research on cognitive dissonance challenged the view of humans as rational decision-makers by showing that people would sometimes perform acrobatic contortions of self-justification rather than accept threatening inconsistencies (Festinger, 1957). Milgram's (1963) studies of obedience revealed how chillingly far people would go in response to simple commands by a person wearing a lab coat. More recently, social psychologists have told us that feeling in control of our actions doesn't make it so, and that much of our behavior can rattle off pretty well without us (Bargh & Chartrand, 1999; Wegner, 2002). After decades of optimism about the steady decline of prejudice, they've told us that rumors of racism's demise have been greatly exaggerated and that valuing equality doesn't guarantee a conscience clear of bias (Devine, 1989; Greenwald & Banaji, 1995). And, if you're thinking that you can't recall a single instance in which these findings would apply to you, social psychologists are here to tell you that's exactly what they would expect someone in your position to say (Pronin, Gilovich, & Ross, 2004).

The excitement is understandable. Ideas that change the way we think about being human are the signposts that let us know psychological science is getting somewhere. Still, that progress leaves a lot of misfortune on the side of the road. These kinds of findings raise important questions about agency, responsibility, and self-control. In this essay we focus on the questions raised by research on implicit social cognition, especially concerning stereotypes and

group attitudes. A common worry is that if biases reside outside of consciousness, then even enlightened and well-intentioned people will be hard pressed to control them. “Never forget,” warned Baudelaire, “when you hear the progress of enlightenment vaunted, that the devil’s best trick is to persuade you that he doesn’t exist,” (1869/1997). How can people control these devilish biases if they don’t know they exist? Does implicit social cognition research show that people hold attitudes and stereotypes of which they are unconscious? And does it cause problems for practical issues of self-control or moral issues of responsibility?

A careful look at the research literature on implicit cognition shows good evidence for some versions of unconsciousness but not others. We distinguish unconscious *influences* of attitudes and beliefs from attitudes and beliefs that are themselves hidden from consciousness, because the literature has very different things to say about them. For instance, it seems clear that people can be, and often are, influenced by attitudes and beliefs without being aware of it. But hard evidence that people have attitudes and beliefs that they don’t know about, or can’t know about when they try, is difficult to find.

At this point, some readers will breathe a sigh of relief because if the research doesn’t prove that people have widespread unconscious prejudices, then maybe it doesn’t threaten comfortable ideas about self-control and responsibility after all. Other readers will feel a little disappointed at the news because if the research doesn’t prove that people have widespread unconscious prejudices, then maybe the science isn’t as groundbreaking as was once thought. But it may be too soon for sighs of relief or disappointment. The varieties of unconsciousness that are well-supported by research pose plenty of problems for people trying to keep their biases under control. In our view, both unconscious attitudes and unconscious influences pose potential problems for controlling actions; but the evidence for unconscious influence is much more substantial than for unconscious attitudes. And so, a lot of people are rubbernecking in the wrong direction, because it is unconscious *influence* that really threatens our ideas about responsibility and self-control.

Does research support the link between implicit measures and unconscious beliefs?

The development of indirect techniques for measuring attitudes about race and stereotypes has been driven by a remarkable surge of interest in implicit social cognition (Greenwald & Banaji, 1995; Petty, Fazio, & Brinol, in press). A common interpretation of these measures is that they reveal unconscious attitudes and beliefs—mental content that people are not aware they posses. This interpretation is so widespread that it is rarely justified or empirically tested. Instead, when an implicit test is used to measure some concept, not only the concept itself, but almost anything associated with that test is often assumed to be unconscious. This interpretation, however, has not gone without criticism. Several authors have noted that the evidence that these tests reveal unconscious attitudes and stereotypes is weak (e.g., Fazio & Olson, 2003; Gawronski, Hofmann, & Wilbur, 2006). Nevertheless, when implicit or indirect tests show different patterns from explicit or direct measures, it is typical to read explanations based on unconscious versus conscious attitudes. In this chapter, we draw on these previous papers as well as on observations from our own research to critically evaluate notions of unconscious attitudes and unconscious influence.

Why do implicit and explicit attitude tests diverge?

The key evidence used to support the argument that implicit tests reveal unconscious attitudes comes from dissociations between implicit and explicit tests. That is, when the results of implicit and explicit measures of the same construct fail to correlate, the conscious-unconscious distinction is immediately invoked. One of the most informative findings comes from a meta-analysis conducted by Hofmann, Gawronski, Gschwendner, Le, and Schmitt (2005). Looking at 126 studies in which participants' attitudes were measured by the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) and via self-report, they found that the average correlation between implicit and explicit attitudes was $r = .24$. Reviews of implicit cognition research using affective priming and other methods have found similarly low correlations (e.g., Fazio & Olson, 2003).

These patterns of dissociation are especially common in studies of racial attitudes, to which we now turn our focus. The most common interpretation is that people report their consciously-held, egalitarian attitudes towards racial minorities on direct measures, while indirect measures reveal the less positive racial attitudes that linger in their unconscious. In essence, explicit measures reveal a vaunted enlightenment, but implicit measures don't fall for the mind's devilish tricks.

The lack of correspondence between implicit and explicit tests, however, may result from other factors that are unrelated to unconscious attitudes. For instance, research subjects may be motivated to present themselves in a positive, unbiased light. Concerns with self-presentation, therefore, may influence the degree to which participants provide honest responses on self-report measures of racial bias, and this may ultimately attenuate the relationship between implicit and explicit measures. Another factor driving the weak correlation may be the low reliability of some indirect measures. Furthermore, when subjects are asked to report their attitudes or feelings, they may be perfectly aware that they tend to get squeamish around certain sorts of people, but they might not feel as though that squeamishness reflects their true attitude. Instead, they may consider it something else—a bad habit left over from their less enlightened days, or a primitive gut reaction not worthy of the title of "true attitude"—a designation reserved for a more enlightened set of beliefs suitable for the polite company of an academic laboratory. Finally, methodological differences between implicit and explicit tests that encourage different mental processes—what we refer to as a lack of 'structural fit'—may also mask higher implicit-explicit correlations. We next look more carefully at each of these issues.

The Motive to Self-Present

Tolerance for the expression of discriminatory and racially-insensitive attitudes has been diminishing. As a result, many people are reluctant to report their true feelings about other races when doing so may paint them in a prejudicial light. The correlation between the implicit and explicit racial attitudes these people report should be low to the extent that: (1) people

deliberately report racial attitudes that deviate from their true beliefs on explicit measures, and (2) they are unable to control the responses reflecting their true racial attitudes on implicit measures. People who are less concerned about appearing prejudiced, on the other hand, should have no reason to alter their responses on explicit measures of racial attitudes. For these people, there should be a much stronger positive correlation between implicit and explicit racial attitudes.

Research has consistently shown that people's willingness to report their genuine attitudes depends on how concerned they are about acting without prejudice. Fazio et al. (1995) examined the relationship between the racial attitudes participants explicitly reported on the Modern Racism Scale (McConahay, 1986), a self-report measure of opinions on racially-charged political issues, and the attitudes revealed using a priming measure of participants' affective associations with Black versus White faces. They found that the strength of the correlation between the implicit and explicit measures differed depending on how motivated participants were to control potentially prejudiced responses. Participants who were highly motivated to suppress reactions that might reflect racial bias on their part showed the typical dissociation between directly and indirectly measured racial attitudes. The racial attitudes they reported on the Modern Racism Scale were weakly (negatively) correlated with the amount of racial bias they demonstrated in their affective reactions to Black and White faces. For participants who were low in the motivation to control prejudiced responses, a stronger positive correlation between implicit and explicit measures emerged.

We obtained similar results in our own research on racially-biased perceptual errors (Payne, 2001). Imagine performing the following task: An image appears on a computer screen in front of you and your job is to identify whether the image is a picture of a gun or a tool. Just prior to the appearance of the gun or tool, you see a photograph of either a Black or a White person's face on the computer. You are instructed to ignore the photograph of the face and to classify the image of the gun or tool as quickly and as accurately as you can. The presentation of

either a Black or White face, followed by the appearance of either a gun or a tool, appears a number of times in succession, and your job, on each trial, is always to correctly identify the guns and tools. We've asked participants to do just this and have found a consistent pattern of results. Participants are more likely to misidentify a harmless tool as a gun when primed with a Black face than a White face (Payne, 2001). The stronger the association people have between 'Black' and 'gun,' the more pronounced is the bias. The weapon identification task can thus be used as an implicit measure of people's racially-biased attitudes towards Blacks.

Paralleling the findings of Fazio et al. (1995), we've discovered that the strength of the correlation between bias on the weapon identification task and participants' self-reported racial attitudes varies based on their motivation to control prejudiced responses. In one study (Payne, 2001), we had participants complete the Modern Racism Scale before performing the weapon identification task. We also measured participants' motivation to control their prejudice using Dunton and Fazio's (1997) Motivation to Control Prejudiced Reactions Scale. The degree of correspondence between race bias on the weapon identification task and the racial attitudes reported on the Modern Racism Scale depended on participants' motivation to control prejudice. For participants low in the motivation to control prejudiced responses, there was a positive relationship between implicit and explicit racial bias against Blacks. Greater implicit bias on the weapon identification task was associated with more negative attitudes towards Blacks on the Modern Racism Scale. The relationship between implicitly and explicitly measured bias evaporated for participants reporting a stronger motivation to control prejudiced responses. That is, participants' responses on the Modern Racism Scale were uncorrelated with their performance on the weapon identification task.

Another line of support for the role of self-presentation motives comes from a study by Nier (2005). This study assessed participants' racial attitudes with a direct and an indirect measure. Using the 'bogus pipeline' manipulation, Nier led half of his participants to believe that attempts at misrepresenting their true racial attitudes would be detected by the experimenter.

This procedure reduced the motive to present oneself in a positive light, as any deviations from the truth would ultimately be exposed. Results mirrored those for motivations to control prejudice. For participants who could act on a self-presentation motive without fear of being detected, the correlation between implicit and explicit racial attitudes was nonsignficant. But for participants who could not act on this motive, there was a significant positive relationship between the implicit and explicit measures of racial attitudes.

Using another approach to investigate motivational concerns, we've obtained comparable results. We've used an indirect attitude measure we call the affect misattribution procedure, or AMP, for short, to assess implicit racial bias (Payne, Cheng, Govorun, & Stewart, 2005). The AMP is based on the idea that people's evaluation of neutral, ambiguous stimuli can reveal something about their underlying attitudes. In a typical study using the AMP, participants view pairs of pictures on a computer screen. The first picture in each pair is an affective prime—an image that invokes a pleasant or an unpleasant reaction, depending on one's attitude. The second image that appears on the computer screen is a Chinese symbol, or pictograph, which participants are asked to rate. These pictographs are unfamiliar to our U.S. participants and are pretested to be neutral in valence. The key dependent measure is participants' ratings of the pleasantness of the Chinese pictographs. The logic behind the AMP is that participants' reactions to the affective primes will be misattributed to their evaluations of the neutral Chinese pictographs. Participants' evaluations of the pictographs can thus provide an indirect measure of participants' emotional responses to the primes.

We have found that the strength of the correlation between self-reported racial attitudes and race bias measured by the AMP varies depending on participants' self-presentational concerns. For example, we recruited a sample of Black and White college students and asked them to report how "warm and favorable" versus "cold and unfavorable" they felt towards Blacks and Whites. This served as our explicit measure of racial attitudes. We used two measures, Plant and Devine's (1998) Internal and External Motivation to Respond Without Prejudice Scale and

Dunton and Fazio's (1997) Motivation to Control Prejudiced Responses Scale, to assess participants' motivational concerns. Finally, we administered a version of the AMP that contained photographs of the faces of young Black and White males as the affective primes to measure implicit racial attitudes.

Both the implicit and explicit measures of racial attitudes revealed a significant own-race preference. White participants evaluated the pictographs paired with White faces more favorably than the pictographs paired with Black faces, and provided explicit reports of more warm and favorable feelings towards Whites than Blacks. Black participants did just the opposite. They evaluated the pictographs paired with Black faces more favorably than those paired with White faces, and their feelings towards Blacks were warmer and more favorable than their feelings towards Whites. Relevant to our present discussion, the implicit-explicit correlation was moderated by participants' general motivation to control prejudiced responses. There was a strong positive correlation between the implicit and explicit racial attitudes of participants who were low in the motivation to control their prejudice. However, the implicit and explicit racial attitudes of participants who were high in the motivation to control prejudice were virtually uncorrelated (Payne et al., 2005).

These patterns suggest that when subjects complete an explicit measure of their racial attitudes, many are faced with a self-control dilemma. When motivations to be unprejudiced clash with automatic affective responses, they have to find some way to resolve that conflict. Research suggests that the discrepancy between motivations and affective responses may itself initiate the conflict resolution process. For instance, Moskowitz and colleagues (1999) suggested that for subjects motivated to control prejudice, processing stereotype-related information activates not only prejudiced responses, but also egalitarian goals. This finding is related to a more general phenomena, in which tempting stimuli elicit not only automatic or impulsive inclinations, but also counteractive control motives (Trope & Fishbach, 2000; see also Amodio

et al., 2004). Because responses on explicit tests are easier to monitor and control than implicit tests, these control efforts can be expected to mainly influence self-reports on explicit tests.

Taken together, these findings indicate that motivational concerns are a major factor determining the relationship between directly and indirectly measured racial attitudes (see also Akrami & Ekehammer, 2005; Dunton & Fazio, 1997; Gawronski, Geschke, & Banse, 2003; Hofmann, Gschwendner, & Schmitt, 2005; Nosek, 2005). The more the lack of correspondence between implicit and explicit tests can be explained by self-presentation motives, the less reason there is to assume that biases driving responses on indirect measures are hidden from consciousness. People can and do report their automatic racial attitudes when they are properly motivated to do so.

Reliability of Indirect Measures

The low correlation between direct and indirect measures of racial attitudes may also be a consequence of the low internal reliability of many indirect measures, particularly measures that use reaction times as the key dependent variable. Compared to direct attitude measures, indirect measures that use reaction times tend to be higher in measurement error (Cunningham, Preacher, & Banaji, 2001). This is not surprising, given the multiple factors that can affect response times on even simple tasks, and the high variability inherent in response time measures (Lane, Banaji, Nosek, & Greenwald, 2007). The measurement error resulting from the use of indirect measures may attenuate the correlation between direct and indirect tests because reliability sets the upper limit on correlations. The highest possible correlation between two tests depends on the product of their reliabilities. Two tests each with reliabilities of .50, for example, cannot correlate higher than .25. Cunningham, Preacher, and Banaji (2001) used latent variable analysis to control for measurement error in implicit tests. In doing so, they obtained a much higher correlation between implicit and explicit racial attitude tests than has typically been found with other statistical procedures.

One way to control for measurement error is to use low-reliability tests, estimate the amount of measurement error, and then estimate what the correlations between tests would be in the absence of measurement error using sophisticated statistical procedures. The success of this approach hinges on the ability to accurately estimate the amount of measurement error. It requires multiple tests for each construct and relatively large sample sizes. Another way to surmount the problem of measurement error is to use indirect measures with higher reliability. The AMP, which does not rely on reaction times, circumvents the issue of high measurement error. Across several studies, the AMP has demonstrated sufficiently high reliability, with an internal consistency that is usually above .80. As a result, we've been able to detect strong positive correlations between implicit and explicit racial attitudes when using the AMP as our indirect measure.

Self-presentation and reliability have been discussed by several authors as reasons for divergence between implicit and explicit tests. To these well-established factors, we add two more that both have to do with what the scores on implicit attitude tests really mean. The first issue is how research subjects construe what their "attitude" is. The second has to do with how well our measurement techniques capture the most important distinctions between implicit and explicit tests.

What is your real attitude, anyway?

Any attitude test is an attempt to operationally define a particular attitude. That includes identifying where a person stands on some dimension, such as from "like" to "dislike," but it also includes some assumptions about what an attitude is to begin with. Some attitude tests, such as a "feeling thermometer," allow subjects to define their attitudes largely for themselves by simply asking them about their feelings. Other attitude tests, such as the Modern Racism Scale, define the attitude for participants by asking about specific policy views, such as whether or not they support busing policies as a means for achieving racial integration. The assumption behind the measure is that subjects' racial attitudes can be inferred from those policy preferences. Implicit

attitude tests also include some assumptions, often tacit, about what an attitude is. Some of those assumptions have been controversial, as researchers disagree about whether the kinds of behaviors measured by particular implicit tests should be considered indicators of attitudes or something else (Arkes & Tetlock, 2004; Blanton & Jaccard, 2006).

If attitude researchers have differing ideas about what an attitude is and whether any given test captures it, it is a good bet that ordinary attitude-holders have an even more diverse set of ideas about what constitutes their attitudes. Oftentimes when participants (and researchers) take an implicit test, they are surprised by what the test reveals about their beliefs and preferences. One interpretation of this feeling is that the attitude is unconscious. If I am surprised by my own test results, then maybe I don't know what my attitude is. From this perspective, taking an implicit test can be like finding out from an x-ray that you swallowed a marble: "How did that get in there?!" Another interpretation of the surprise is that people may not consider the associations, feelings, or experiences driving their responses on implicit tests to reflect their "real attitude." In this case, the first response is "Oh no -- a foreign object?!" but then, "Oh, wait, you just mean that marble I swallowed last Tuesday." The trick is knowing when the researcher and the subject are talking about the same marbles.

An intriguing study has recently made some headway on this problem. Ranganath, Smith, and Nosek (2008) gave research subjects three reaction-time-based implicit measures of attitudes toward gay people. They also asked subjects to rate their attitudes toward gay people using scales that distinguished between "gut reactions" and "actual feelings." The gut reactions subjects reported were more negative than their actual feelings. And importantly, those gut reactions corresponded well with the implicit measures, whereas ratings of actual feelings did not. This pattern suggests that when studies of stereotypes and intergroup bias ask people to report their attitudes, the low correlations frequently found with implicit measures may stem in part from the fact that subjects don't consider their "gut reactions" to be their "actual feelings." It also suggests that subjects have some awareness of the attitudes revealed by implicit tests

because when asked the "right" questions, they can report them in a way that matches their responses on implicit tests.

This study illustrates the importance of carefully considering what subjects think words like "attitudes" and "feelings" mean. It is difficult to know what to make of results when the researcher and the subject don't think they mean the same thing. The more general version of this argument is that when measuring implicit and explicit responses, it is important for the tests to be controlled in such a way that the only key difference between them is along the dimension of interest—be it consciousness versus unconsciousness, automatic versus controlled responses, etc. If the key difference is confounded with other inconsistencies between the two tests, such as differing meanings of "attitude" or different construals of the attitude object, then it is hard to know what the results mean. Below we discuss some of the problems with using implicit and explicit tests that differ in ways beyond the implicit-explicit distinction and describe a potential solution.

Structural fit: Equating implicit and explicit tests on extraneous differences

The obvious difference between implicit and explicit tests is that implicit tests are supposed to be implicit–that is, they're supposed to measure automatic or unconscious processes—and explicit tests are supposed to be explicit measures of consciously-controlled processes. But take a minute to think about all of the other ways that implicit and explicit tests tend to differ. In other words, the weak correlation between implicit and explicit racial attitudes may often be driven by differences in the way direct and indirect measures are typically structured. Take, for example, the low correlation that many researchers have found between racial attitudes reported on the Modern Racism Scale (MRS) and performance on the IAT. The difference between these measures that is usually of greatest theoretical interest is that one test, the MRS, asks participants to directly report their attitudes, whereas the other test, the IAT, provides an indirect index of participants' racial attitudes. However, these two measures differ in a number of other ways that are entirely independent of the direct-indirect (or, implicit-

explicit) distinction. For instance, the MRS asks participants to indicate how much they agree or disagree with verbal statements that capture attitudes about racially-charged political issues. Completing the MRS requires that participants read complex propositional statements and translate their attitude into a response that falls along a bipolar (disagree-agree) continuum. Participants taking the IAT are asked to indicate whether words or pictures representing different categories are 'good' or 'bad' based on a decision rule that is established at the beginning of the test. As this comparison illustrates, although the Modern Racism Scale and the IAT differ with respect to how directly they measure racial bias, they also differ in ways that are unrelated to the implicit-explicit distinction. Without controlling for these incidental differences, conclusions about the underlying cause of the low correlation cannot necessarily be drawn.

Using the AMP as an indirect measure of racial attitudes, we have found that the strength of its relationship with direct attitude measures depends heavily on how closely matched the structural components of the two tests are. In a recent series of studies, subjects performed the AMP, with photos of Black and White male faces as the affective primes that preceded the presentation of the neutral Chinese pictographs. Participants were instructed to ignore the pictures of the faces, and to rate the pleasantness of each Chinese pictograph on a 4-point scale. In a second version of the AMP, subjects saw the same face-pictograph pairs, but were instructed to ignore the Chinese pictographs and rate the pleasantness of the Black and White faces on the same 4-point scale. The standard version of the AMP, in which participants rated the Chinese pictographs, served as an implicit or indirect measure of participants' attitudes towards Blacks and Whites. The modified version, in which participants provided their direct evaluations of the Black and White faces, served as an explicit measure. But unlike other studies comparing implicit and explicit tests, these measures were equated on their other structural features. Subjects were presented with the same items, given matched tasks, and responded using the same scales. Correlations between implicit and explicit attitudes were

strong using this method, in contrast to the many weak correlations that have been reported comparing reaction time measures to self-report tests.

In another study, we examined the relationships between several different implicit and explicit measures of racial attitudes that varied in their degree of structural fit. We included two versions of the AMP – one that used Black and White faces as primes as described above, and one that used verbal group labels (i.e., African Americans, Blacks, European Americans, Whites). Subjects rated the pictographs in some blocks, and they rated the primes in other blocks. In addition to these implicit and explicit tests, subjects completed three traditional self-report measures of racial attitudes, including a "feeling thermometer," the Modern Racism Scale, and the Attitudes Toward Blacks Scale (Brigham, 1993). To compare how well these tests cohered, we first rank ordered them by how well their structural features were matched. As summarized in Figure 1, we found that the magnitude of the correlation between each of the direct measures and scores on the indirect AMP varied as a function of structural fit. Figure 1 shows the strength of the relationship between structural fit (reverse rank ordered, so that higher numbers reflect greater structural fit) and the size of the implicit-explicit correlation. Structural fit was correlated at a strikingly high .90 with the size of implicit-explicit correlations. When the tests had poorly matched structures, as in virtually all implicit attitude studies, the implicit-explicit correlations were very low. When the tests were structurally equated, implicit and explicit responses were strongly positively correlated, painting a much more consistent picture of people's attitudes about race.

Summary

It appears that the verdict on unconscious race bias and the existence of attitudes that are hidden from consciousness altogether may have been issued to soon. With the development and growing popularity of implicit attitude measures came the assumption that these instruments were tapping into sentiments that people harbor unknowingly. But closer inspection of the key evidence supporting this claim—the disconnect between the results of

implicit and explicit tests—reveals a number of factors unrelated to the conscious/unconscious distinction that help to explain this finding (see Gawronski et al., 2006, for a similar conclusion). Self-presentational concerns may lead some participants to alter their responses on self-report measures of prejudice in order to portray themselves in a more egalitarian light. Furthermore, differences in the indicators that researchers and participants are using to determine what a 'true attitude' is, the failure to equate implicit and explicit tests that are being compared on all but one critical dimension, and the low reliability of implicit measures that rely on response latencies can also attenuate the correlation between implicit and explicit attitudes about race. In the section above, we reviewed research demonstrating that when each of these factors is taken into account, the correspondence between implicit and explicit attitudes increases substantially. The stronger correlation has important implications for the hidden bias debate: It suggests that people do, in fact, have access to the mental content they reveal on implicit tests.

If unconscious attitudes about race—previously seen as the major obstacle for controlling one's prejudice—aren't as widespread as was once thought, where does this leave people who are worried about keeping their biases in check? Still in plenty of trouble, we argue. A major assumption has been that consciously-held attitudes are subject to voluntary control. That is, to the extent that we are aware of our feelings towards others, we should also be aware of how these feelings shade our interactions. As a result, we are (and should be held) responsible for the behavior we perform that stems from these attitudes. The flipside of this logic is that unconscious attitudes shade our interactions in ways that we can not control. It's difficult to control the impact of feelings we don't even know we have. The idea of unconscious attitudes has thus given rise to considerable debate concerning how accountable we should be for the behaviors they produce.

We argue that an attitude need not be unconscious in order to influence our thoughts and behaviors without our awareness. There is plenty of research indicating that attitudes of

which we are perfectly conscious can and do influence our thoughts and behaviors in countless *un*conscious ways. In the section below, we briefly review some of the seminal research on unconscious influence. We then turn to several findings from our own research that highlight the unconscious influence of attitudes about race. As our work demonstrates, we may be painfully aware of our prejudices towards others, but blissfully *un*aware, nonetheless, of how these prejudices come to affect our judgments, preferences, and behavioral decisions. And as it turns out, unconscious influence may be every bit as fascinating or threatening, depending on how you view it.

Unconscious Influence of Attitudes

The notion of unconscious influence is hardly new to psychology. The story of Clever Hans should have a familiar ring for anyone with at least a semester of general psychology under his or her belt. Clever Hans was a student of Wilhelm von Osten, a German school teacher during the early 1900s. Under the tutelage of von Osten, Hans had acquired an array of skills: he could multiply fractions, perform long division, construct grammatically-correct sentences, and even identify the tones in a musical scale. What had earned Clever Hans his nickname was the fact that he was a horse. Von Osten would present Hans with a question or problem to solve, and Hans would respond by stomping his hoof a certain number of times (Pfungst, 1911).

The pair drew audiences from throughout Germany. For some, witnessing Clever Hans perform his feats was enough to convince them that he was a remarkable horse. Others remained skeptical, questioning just how clever Hans really was. Detractors assumed that Hans' abilities were part of an elaborate hoax orchestrated by von Osten. In response to mounting controversy, a team of scientists set out to investigate whether Clever Hans was, in fact, a sham. The scientists discovered that Hans' hoof stamps were driven by subtle, nonverbal cues that the person questioning him would emit in anticipation of the correct answer. These cues appeared to be unintentional and involuntary—even the team of scientists had emitted them in the early

stages of their investigation. Much to von Osten's dismay, Clever Hans' extraordinary skill was not the ability to read and do math, but the ability to pick up on nonverbal cues.

We tell the story of Clever Hans because it demonstrates how our attitudes—and expectations—can leak out without our awareness. More contemporary research by Rosenthal and Jacobson (1966; 1968) showed that it was not only horses that could pick up on these cues, but that children could, too. In a now-classic study, the researchers informed a sample of kindergarten teachers that several of their pupils would 'bloom' intellectually throughout the coming year. Although they had been chosen at random, the children who were identified as potential 'bloomers' did, in fact, demonstrate superior academic ability by year's end. The schoolteachers had been emitting subtle cues based on their expectations that the bloomers would outperform their less gifted peers, and it was the teachers' expectations of success that caused the actual surge in the bloomers' intellectual growth. At one point dubbed the "Clever Hans effect," this teacher expectation finding later became known as the "Pygmalion effect," perhaps because it's more polite to compare schoolchildren to a Greek statue than to a horse.

Other seminal findings on unconscious influence come from research by Nisbett and Wilson (Nisbett & Wilson, 1977; Wilson & Nisbett, 1978) demonstrating that people lack awareness of their higher order cognitive processes. That is to say, people may know exactly *how* they feel about various stimuli (e.g., objects, people, social groups), but their explanations for *why* they feel a certain way aren't based on any actual introspection into their thought processes. Instead, people generate explanations for their feelings and decisions based on *a priori* theories that seem plausible (e.g., choosing a college because it is nationally ranked) or salient features of their environment (e.g., attributing a bad mood to rainy weather). In one of their most well-known illustrations of this phenomenon, Nisbett and Wilson (1978) placed four identical pairs of stockings on a table in a department store and asked shoppers passing by to identify the pair of the highest quality. Despite the fact that the stockings were virtually identical, the shoppers' preferences showed evidence of a strong positioning effect. The further

to the right the stockings appeared in the line-up, the greater the likelihood that shoppers identified it as the best in quality, with 40% of shoppers choosing the pair positioned on the far right. When asked to explain how they had reached their decision, no shoppers indicated that they had been influenced by the placement of the stockings on the display. Rather, they cited differences in knit, weave, sheerness, and elasticity—explanations that seemed plausible, but reflected a lack of awareness of the actual mental processes that had led to their decision.

A related notion is that of mental contamination, a term Wilson and Brekke (1994) have used to describe unwanted influences on people's thoughts, emotions, and behaviors. According to Wilson and Brekke, mental contamination can occur for two reasons: people are unaware of the unwanted influence altogether; or, they are aware that the influence may be taking place, but are unable to control it. For example, employers may want to hire the best person for the job, but they may not realize how an applicant's race, gender, or degree of physical attractiveness may be affecting their decision. Alternatively, employers may be aware of the potential for such factors to shade their evaluations of an applicant, but may still be unable to prevent the "contamination" from taking place. Research from our own lab suggests that in either case, unconscious influence seems to be the source of the problem.

Knowing when we are biased

Do people lack insight into the mental processes that result in patterns of racial bias? One way that we have addressed this question has been to look at the correlation between people's subjective experience of bias—whether or not they think they have been influenced by racial cues—and the degree to which they actually demonstrate it. If people are aware of the influence of their attitudes about race on their judgments, evaluations, and behavior, we would expect to find a positive relationship between people's perception and actual demonstration of bias.

We investigated this relationship in a study in which participants performed a simple memory task (Payne, Jacoby & Lambert, 2003). We gave participants a list of names that were

stereotypical of either Black or White males (e.g., Jamal versus Gregory). Each name on the list was paired with one of two occupations: basketball player (an occupation that is more stereotypical for Black males) or politician (an occupation that is more stereotypical for White males). Thus, participants memorized a list of name-occupation pairs, some of which were consistent with prevailing stereotypes and some that were not. Later, we asked participants to do two things: recall the occupation that had been paired with each name and indicate how confident they were that each of their answers was correct.

This memory task allowed us to separate the automatic and controlled processes that were driving participants' responses. Once again, imagine yourself as a participant in this study. You are prompted with the name 'Keyshawn' and asked to remember whether you'd seen this name paired with the occupation 'politician' or 'basketball player' a few minutes back. One of two things may happen. You may have a conscious recollection of the occupation that was actually paired with this name. In this case, the answer you choose based on your recollection results from the controlled use of memory and will in most instances be correct. However, when prompted with the name Keyshawn, the stereotype-congruent occupation may also come to mind automatically. This automatic bias may occur regardless of whether the stereotypical occupation is the correct or incorrect response.

We were interested in how both of these processes might relate to participants' subjective experience. We used participants' confidence in their responses as an indicator of the degree of insight they had into the accuracy of their memory and their degree of automatic bias. Although participants showed good insight into their ability to recall the accurate profession overall, as evidenced by the positive correlation between confidence and accuracy, they showed poor insight into their degree of automatic bias. That is, their degree of confidence and demonstration of bias were virtually unrelated. Another way of looking at this is that when participants were correct, they knew they were correct. When participants were wrong, they were more likely to misremember that stereotypically White names had been paired with

'politician' and that stereotypically Black names had been paired with 'basketball player' than to misremember the opposite pairing. And, they were just as likely to show this bias when they reported perfect confidence in their memories as when they reported having no confidence at all.

We have observed a similar effect using both the weapon identification task and the AMP. In a study using the weapon identification task, we asked participants to indicate how accurate they felt they were in their classification of the weapons and household tools (Payne, Lambert & Jacoby, 2002). Recall that in this paradigm, participants are asked to classify images they view on a computer screen as either a 'gun' or a 'tool' and that these images are paired with a photograph of either a Black or White male face. This time, however, we varied the amount of time participants had to make the classifications, with deadlines ranging from 200 to 700 milliseconds after the presentation of the photos. The variation in deadlines allowed us to examine more closely the automatic and controlled processes driving participants' responses. With a longer deadline, participants would have enough time to discriminate guns from tools on the basis of the objective features of each image. In other words, they would be able to think carefully about whether they were viewing a picture of a gun or a tool. Participants' ability to exert cognitive control in these instances would result in a high rate of accurate responding. With a much shorter deadline, participants would have to rely, at least in part, on automatic processing. In these instances, participants' responses would be much more likely to reflect an accessibility bias. To the extent that participants associate guns more strongly with Black versus White males, they should show a greater tendency to make stereotype-congruent identification errors.

As predicted, accuracy rates were highest when participants were given 700 ms to respond and lowest when they had only 200 ms. When participants did make errors, their errors were consistent with prevailing stereotypes. They were more likely to misidentify tools as guns after being primed with a Black face, and were more likely to misidentify guns as tools after

being primed with a White face. Participants' confidence in their accuracy and their actual rate of accuracy were positively correlated, but this was driven entirely by the positive correlation between participants' confidence and cognitive control. There was no relationship between confidence and the amount of automatic bias they demonstrated.

These findings highlight a major obstacle we encounter when we try to control prejudice. It's hard to know how to control unwanted influence if we can't tell when it's at work. In support of this, we've investigated two specific strategies for reducing bias: (1) warning people about the potential for bias; and (2) giving people the opportunity to refrain from making potentially-biased responses altogether. We've discovered that both strategies are ineffective precisely because they rely on people's ability to detect bias as it is happening. Consider a study in which participants performing the weapon identification task were randomly assigned to one of three conditions (Payne, Lambert & Jacoby, 2002). In the first condition, participants were told to try their best not to let the pictures of the Black and White faces influence their identification of the weapons or tools. In the second condition, participants were actually encouraged to stereotype, and were instructed to use the Black and White faces as cues for helping them identify the weapons and tools. In the third condition, participants were given no explicit instructions regarding the pictures of the Black and White faces, and thus served as a 'no goal' control. We found that participants' who were instructed to use race as a cue did, in fact, report significantly stronger intentions to do so than participants in the other two conditions. In spite of this, the amount of bias that participants actually did demonstrate did not vary across conditions. Participants mistook tools for guns with greater frequency after being primed with a Black face and mistook guns for tools with greater frequency after being primed with a White face, regardless of whether they had been trying to use or refrain from using race as a cue. It appeared that giving participants any explicit warning about race, whether it was to use or to avoid using race as a cue, seemed to activate stereotypes about race and increase the magnitude of race bias.

A similar finding has emerged in our research using the AMP. That the AMP provides evidence of unconscious influence should come as no surprise. The success of the procedure as a measure of attitudes relies on the unintentional transfer of people's affective reactions to various primes. In a sense, it's a measure that is defined by unconscious influence. Interestingly, this misattribution of affect occurs even after participants are told that the primes may be influencing their evaluations of the neutral Chinese pictographs. Participants who are warned about the potential for bias and instructed not to let their reactions to pleasant and unpleasant photos (Payne et al., 2005; Experiments 1 & 2) or Black and White faces (Payne et al., 2005; Experiment 5) influence their ratings of the pictographs show just as much bias as participants who are not.

We've also looked at another strategy for controlling bias. Imagine yourself as a contestant on your favorite game show. The answers you provide to various questions (or the questions you provide to various answers, if *Jeopardy!* was the show that came to mind) will fall into one of two categories: correct and incorrect. But what if you could refrain from providing an answer altogether if you weren't sure your answer was correct? Using the paradigms we've discussed, we've given participants an opportunity to 'take a pass' when they feel that their responses may be biased. For instance, in a study high in external validity for anyone who's ever forgotten an acquaintance's name at a social gathering and faced the dilemma of whether to hazard a guess or avoid incrimination by keeping one's mouth shut, we had participants memorize a list of names and professions that were stereotypically Black or White, but this time gave participants the opportunity to 'pass' on certain trials if they were unsure of the correct answer (Payne, Jacoby, & Lambert, 2003). We wanted to know whether people would hold their tongues, so to speak, when their recall of the correct answer was shaky. To examine this, we varied the format of the items on the recall test so that each participant received a combination of forced and free response trials. On the forced response trials, participants had to provide an answer. If they had trouble remembering the correct occupation for that name, they had to make

a guess. On the free response trials, participants had the option to 'pass' if they couldn't recall the correct occupation. Thus, participants could reduce the number of errors they made on the free response trials by cutting out the guess-work. We found that providing participants with the 'free response' option did little to reduce their stereotype-congruent errors in recall. Because participants did not know *when* their responses were being influenced by stereotypes about race, the opportunity to hold their tongues did not afford them better control over their bias.

Yet another challenge that unconscious influence poses for the control of prejudice comes from research by Irene Blair, Jennifer Eberhardt, and their colleagues (Blair, Judd & Chapleau, 2004; Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006) showing that despite our best attempts to curtail unwanted influence, our attitudes about race may bias our judgments in new and unexpected ways. Their work suggests that controlling the influence of attitudes about race may be like using a bucket to stop rain as it falls through a leaky roof—placing the bucket under one hole doesn't prevent water from dripping down through another. The research by Blair and Eberhadt comes in the wake of the acknowledgement that the color of a defendant's skin has an impact on the length and severity of the sentence he or she receives. A number of states have established sentencing guidelines that require legal decisions to be made on the basis of objective criteria, and these guidelines have been enacted in part to address well-documented racial disparities in the criminal justice system (see Tonry, 1995). Even with these guidelines, subtle forms of race bias continue to shape legal outcomes. In their investigation of the criminal sentences assigned to a random sample of inmates in the state of Florida, Blair, Judd, and Chapleau (2004) found that independent of racial group membership, there was a significant relationship between the severity of the punishment received and the degree to which a defendant's facial features were stereotypically Afrocentric. While there were no significant differences in the sentences that Black versus White defendants received once prior criminal records and the severity of the crime were statistically controlled, defendants with more stereotypically Afrocentric facial features (as determined by nose width, lip thickness, and hair

texture) received longer sentences than defendants with less Afrocentric features. This difference remained even after criminal history and the severity of the offense were taken into account. Defendants with more stereotypically Afrocentric features received harsher sentences, regardless of whether they were Black or White.

Eberhart and colleagues (2006) observed a similar effect in cases involving the death penalty. Specifically in capital sentencing cases involving a Black defendant and a White victim, they found that defendants with highly Afrocentric facial features were twice as likely to receive the death penalty as defendants with less Afrocentric features. This disparity held even after aggravating and mitigating circumstances, severity of the murder, and the socioeconomic status of both the victim and defendant were controlled. These studies suggest that while sentencing guidelines can help reduce the likelihood of discrimination based on a defendant's skin color, race-related bias may still influence legal outcomes in less obvious ways.

The difficulty in overcoming unconscious influences of racial bias has parallels in studies of the effects of social context on person perception. Trope (1986) has distinguished between influences of context on identification of behaviors versus influences of context on adjustment of judgments for situational demands. When context is used to adjust judgments, perceivers make deliberate inferences, for example, that fearful actions in a frightening situation do not necessarily mean that the actor is fearful, just that he is responsive to the situation. In contrast to this type of deliberate inference, context also influences people automatically by biasing their interpretations of exactly what the act is to begin with. So a quick movement that happens immediately after a loud noise might be interpreted as a fearful startle, while the same movement without the noise might be interpreted as excitement. Trope and Alfieri (1997) found that influences of context on inferential adjustments depended on processing resources, but influences of identification did not. Instead, influences of context on identification depended on how ambiguous the action was, but not on cognitive resources. The important point for the present purpose is that deliberately thinking about the judgment was not sufficient to overcome

the influences of context on the identification of behavior. The reason is that the process of interpreting the behavior was presumably influenced unconsciously, and only the end product of that process was available to introspection.

These studies illustrate how in some cases, by the time a person tries to inspect their own judgments, it may be too late. In some cases, unconscious influences on the interpretation of ambiguous events may constrain the kinds of explicit inferences that are made. Uhlmann and Cohen (2005; Experiment 1) asked participants to evaluate one of two candidates for the job of police chief. One candidate was presented as being "street smart" but lacking in formal education, whereas the other candidate was presented as being well-educated but lacking in "street smarts." Participants were also led to believe the candidate was either male or female. The key dependent measures were how likely participants would be to hire the candidate as police chief and how important various attributes, including being street smart and being well-educated, were for the job. Uhlmann and Cohen found that participants were more likely to hire the male candidate than the female candidate—a finding that fits with traditional gender stereotypes linking men with both law enforcement and positions of leadership. What's fascinating is that participants' ratings of the importance of being street smart versus being educated fluctuated in such a way that supported their preference for a male candidate (see also Norton, Vandello & Darley, 2004). Participants who evaluated a male candidate with street smarts rated this attribute as more important than educational background, while those who evaluated a male candidate with a formal education rated education as more important than street smarts.

That is to say, they selectively valued or devalued the attributes apparently to justify their preferences for a male candidate. These mental gymnastics allowed participants to maintain a belief in their objectivity. In fact, self-perceived objectivity was actually correlated with actual amount of bias. Those participants who felt their hiring decision was based on "rational,"

"objective," and "logical" decision-making were the ones who showed the most bias in the value they placed on job criteria.

Summary

In this section, we reviewed research demonstrating the pervasiveness of unconscious influence. From intelligent horses to controlling the influence of stereotypes and attitudes about race, unconscious influence has been shown to be difficult to keep at bay. Across several studies, we've found that people aren't particularly good at detecting bias as it is happening. This is evidenced by the striking disconnect between people's subjective experience of bias (whether or not they think they've been biased) and their actual demonstration of it. The inability to detect bias in real time renders certain strategies for controlling prejudice ineffective. For example, warning people about the potential for unwanted bias and providing them with the opportunity to refrain from making potentially-biased judgments do little to reduce the bias from taking place. In some instances, these strategies actually increase the likelihood of bias. In others, bias may be curbed on one level, but it may continue to 'leak out' on other levels. Thus, the unconscious influence of stereotypes and attitudes about race provides a major challenge for self-control.

Strategies for Self-Control

The research we have reviewed thus far has converged on two main points. First, implicit biases can often be reported accurately on explicit measures under the right conditions. The conditions are "right" when a) subjects are motivated to report their attitudes, b) the tests are reliable enough to detect them, c) experimenters and subjects have the same psychological constructs in mind, and d) implicit and explicit tests do not differ in ways that confound the implicit/explicit distinction. These findings suggest that the mental content assessed by implicit tests is not necessarily inaccessible to consciousness. The second point is that even consciously reportable attitudes and beliefs can influence behaviors without our awareness. When people attempt to control their behaviors, such unconscious influence presents just as many challenges.

People aren't particularly good at detecting when they are versus when they are not being biased, and as a result, attempts at controlling one's prejudice may be unsuccessful.

How, then, can prejudice be kept in check? One frequently advocated solution to the problem of hidden bias is consciousness-raising. The idea is that by taking an implicit test, people can discover their unconscious attitudes and begin to address them. This is similar to the logic underlying the classic psychoanalytic notion that by gaining insight into our unconscious mind, we can exercise control over it. But if unconscious attitudes and beliefs are not at the heart of the problem, then consciousness-raising may not be the best solution. Our emphasis in this chapter has been on the pervasiveness of unconscious influence, rather than on attitudes that are themselves hidden from consciousness, and we believe this analysis points to a different set of strategies for self-control. In this final section, we describe some strategies that may be particularly helpful for reducing unconscious influence.

Solution 1: Limiting the potential for bias

Warning people about the potential for bias does little to prevent the bias from occurring. Nevertheless, warnings may provide an indirect benefit for people struggling to control unconscious influence. If people are aware that they are entering situations where the potential for bias is high, they can take measures to shield their mental processes from the unwanted influence ahead of time (Wilson & Brekke, 1994). Consider a hiring practice that many symphony orchestras have adopted over the years. Historically, orchestras in the United States and Europe were comprised almost entirely of men, a trend that was due in large part to negative stereotypes about women's musical ability. To ensure that the very best musicians were being hired for the job, many orchestras implemented the practice of 'blind auditions.' Musicians were required to audition from behind a screen that masked their physical appearance from the judges evaluating their performance. Doing so made it impossible for judges to evaluate them on the basis of anything other than the sound being produced. Indeed, research has shown that as the popularity of this procedure increased, the gender composition

of symphony orchestras began to change. Today, the orchestras that hold blind auditions are the ones with the highest proportion of female musicians (Goldin & Rouse, 2000).

Comparable practices are routinely employed in other areas as well. Blind and double-blind experimental designs are now standard in many scientific fields. Papers submitted to academic journals frequently undergo a process of blind peer review. Teachers often grade assignments blind to students' names to ensure that the evaluations they make are based on merit alone, and not on attitudes they may have about particular students. These measures help eliminate unconscious influence by placing a barrier between the person making the evaluative judgment and all non-essential features of the target. In other words, decision-makers are still making a subjective judgment, but situational constraints have been put in place that make it impossible for the mental processes leading up to the judgment to be influenced by extraneous factors.

There are undoubtedly times when this strategy may be difficult or impossible to employ. It would be hard to prevent jurors from observing the race or gender of a defendant sitting in a courtroom or medical professionals from noticing these characteristics in the patients they treat. In these situations, it may be still be possible to limit the potential for bias by relying on statistical prediction rules. Statistical prediction rules and their application in both legal and medical domains have been part of a larger debate concerning clinical versus actuarial judgment (Dawes, Faust & Meehl, 1989). A judgment that is 'clinical' is based on human reasoning, whereas 'actuarial' judgments are automated and based purely on statistical modeling (such as probability and linear regression models). Actuarial judgments are automated in the sense that a specific piece of information or combination of data will always yield the same result.

In an investigation of over 100 studies directly comparing clinical and actuarial predictions, Paul Meehl (1986; Dawes, Faust, & Meehl, 1989) found that actuarial judgments were consistently more reliable and accurate. In virtually all of the studies, actuarially-based predictions were more accurate than the predictions made by experienced professional in the

relevant fields—even when the people making the predictions were given an informational advantage. Statistical prediction rules, for example, provide more accurate predictions of academic success than admissions officers are able to make (e.g., Dawes, Swets & Monahan, 2000), are better at predicting whether violent criminals will commit future acts of violence (Monahan, 1995), and lead to more accurate diagnosing by mental and physical health practitioners (e.g., Goldberg, 1968). So what are the implications for self-control? If people relinquish their control over the decision-making process and instead rely on statistical models, there will be less opportunity for unconscious influence to interfere and people may ultimately be happier with the results.

As a strategy for combating unconscious influence, putting on literal or statistical blinders may be particularly effective for several reasons. First, it circumvents the stage in mental processing when people may be especially susceptible to influences they aren't aware of. That is not to say that evaluating the quality of a musical performance, deciding whether to accept a manuscript for publication, or choosing what criteria to include in a prediction equation isn't subjective. But people *can* be confident that factors like race and gender that are irrelevant to the evaluation at hand won't be influencing their decisions. The second key element of this strategy is that it is proactive. Because constraints are put into place ahead of time, successful self-control does not hinge on people's ability to detect bias as it is happening. Research in other domains has shown that proactive forms of self-control can be much more effective than strategies that are put into place after problems arise. Aspinwall and Taylor (1997), for example, have argued that implementing proactive strategies for coping with potential stressors in one's life can not only lessen the impact of stress when it does occur, but it can also reduce the likelihood that the stress will occur in the first place. Likewise, strategies for eliminating or at least limiting the potential for unwanted bias that are implemented proactively should help surmount the problem of unconscious influence.

Solution 2: Proactive control over automatic responses

Warning people about the potential for bias, in itself, doesn't prevent unconscious influence from happening. What's more, giving people the explicit goal of not being biased—telling them *not* to use race as a cue—can actually increase the likelihood that the bias will occur. Recall the weapon identification study described earlier (Payne, Lambert & Jacoby, 2002) where we gave participants the explicit goal of not being influenced by race, encouraged the use of race as a cue, or gave participants no explicit goal at all. The participants who'd been told to avoid using race as a cue showed just as much bias as the participants who'd actually been encouraged to rely on racial stereotypes. Similarly, MacCrae, Bodenhausen, Milne, and Jetten (1994) have found that trying not to think in stereotypical terms can backfire, leading to an ironic surge in stereotypical thinking later on. In their research, participants who were given explicit instructions to avoid thinking about a social target in stereotypical terms showed an increase in stereotypical thinking when they stopped trying to suppress. This rebound effect had significant implications: Participants experiencing a post-suppression surge in stereotype-related thoughts were more likely to act on their prejudices in a later phase of the study. Additional support comes from research by Trawalter and Richeson (2005) documenting executive impairment following cross-racial interactions. Building on their finding that people show cognitive impairment after interacting with a member of a different race (particularly when people have a negative attitude about the other race), Trawalter and Richeson examined the impact of different interaction goals. They found that participants who went into a cross-racial interaction with the goal of avoiding the expression of prejudice showed more subsequent cognitive impairment than participants who entered the situation with the goal of having a positive interracial exchange.

Although trying not to be biased may backfire in many contexts, there may be ways to proactively exert control over automatic responses before we are in their grasp. Specifically, we've found that race bias can be reduced by asking participants to commit to implementation intentions that activate counterstereotypical thoughts (Stewart & Payne, in press).

Implementation intentions are plans that link a behavioral opportunity to a specific response. They take the form of "if, then" guidelines that use cues in a person's environment to dictate specific goal-directed behavior (Gollwitzer & Brandstatter, 1997). For example, if you've borrowed this book from the library, you probably have the broad intention to return it on time. An efficient way to help you achieve success would be to devise an implementation intention to carry out once you've finished the book: *if* you see it sitting on your desk, *then* you will put it in your briefcase or backpack so you can drop it off.

In an initial study using implementation intentions, we had participants perform the weapon identification task. Rather than give them the fairly abstract goal to avoid being biased, we randomly assigned them to commit to one of three implementation intentions. In the first condition, we gave participants an implementation intention we believed would override the activation of the stereotype linking Black males with threat. We told them that during the weapon identification task, every time they saw a Black face, they should immediately think 'safe.' We had participants commit to this implementation intention by saying to themselves "whenever I see a Black face on the screen, I will think the word 'safe.'" Participants in the other two conditions were told to think either "quick" or "accurate" whenever a Black face appeared on the screen, and they were also asked to commit to these intentions. These were used as "dummy" intentions because the task already required subjects to respond quickly and accurately. We chose 'think quick' and 'think accurate' so that participants in these control conditions would carry out an implementation intention that was relevant to the task at hand, but unrelated to stereotypes about Blacks.

Participants were more likely to mistake tools for guns when primed with a Black face, and were more likely to mistake guns for tools when primed with a White face. However, this effect was further modified by participants' implementation intention (see Fig 2). Only participants in the 'think quick' and 'think accurate' conditions showed the typical pattern of race bias. Participants in the 'think safe' condition were statistically no more likely to mistake

tools for guns after seeing Black faces than White faces. The implementation intention reduced racial bias even though subjects responded just as quickly as in the other conditions. Moreover, the intention took effect within the first several trials, and remained effective throughout the entire task. Even under conditions of fast and demanding responses, implementation intentions appear to be highly effective at reducing automatic bias.

Solution 3: Creating cognitive connoisseurs

A final approach is to strengthen people's awareness of the influences on their thoughts and behaviors over time. Discussions of consciousness and unconsciousness often have a static tone. A belief or process is said to be conscious or unconscious, without considering that what is unconscious at one moment (e.g., the exact position of your left thumb) may be conscious at another moment (e.g., now). This example hinges only upon directing your focus, but in other cases, bringing mental processes into consciousness may require training and practice. Consider, for example, how experienced meditators learn to attend precisely to their breathing and body position, or how wine connoisseurs learn to become aware of many aspects of wine that they formerly could not discern. The last approach we discuss for reducing unconscious influence is to become a *cognitive connoisseur*. That is, one way to overcome unconscious influences is to learn, with practice, to become aware of subtle influences on thought and behavior that may not have been perceived before.

One way to accomplish this goal may be through meta-cognitive training. The objective of meta-cognitive training is to teach people how to monitor their own thought processes. We know of no specific applications of meta-cognitive training within the context of prejudice reduction, but programs designed to teach people to become better 'thinkers' have been successful in other domains. For example, metacognitive training has been used within the context of computer-based learning to help students develop critical problem-solving skills. In a study by Teong (2003), 11 and 12 year-olds with low scores on a math achievement exam participated in a training program that encouraged them to use metacognition to solve math

word problems. After being presented with a math problem, students were prompted to ask themselves out loud if they understood what the problem was asking. They were also instructed to think of the possible strategies they could use to help solve the problem, to ask themselves if they were on the right track after choosing and implementing a strategy, and, after obtaining a solution, to ask themselves if their answer made sense. Compared to students in a control condition who had been given the same math problems without the metacognitive prompts, students who received the metacognitive training showed greater improvement in their math skills over time.

We believe this study points to a crucial first step that must be taken by aspiring cognitive connoisseurs. That is, the first step in a meta-cognitive training regimen may be as simple as getting people to stop and think about all potential sources of influence before they make a judgment or decision. Think about the countless evaluations you make in a day. Chances are, in only a handful of the cases did you ask yourself why you may have felt or acted a certain way, and, in still fewer cases did you stop to think about potential sources of influence beforehand. When people do stop to consider these things (or are prompted by an experimenter), their analysis probably isn't based on a careful consideration of *all* possible sources of influence, but rather on the sources that seem most plausible or are the most salient. Asking people to do something as simple as listing all possible factors that *could* exert an influence in their heads—even the things they're certain they won't be swayed by—may increase their awareness of influences that might otherwise go undetected.

These considerations point to the direction of attention as an important factor determining what mental contents are conscious. A related approach has been taken in studies of emotion awareness, a topic that could be instructive for understanding awareness of attitudes. Lane's (2008) model of emotional awareness argues that the ability to become aware of emotional processes is a cognitive skill like any other. Emotional awareness, in this view, can occur at different levels of sophistication, from simple and vague feelings (e.g., I feel bad) to

discrete emotions (e.g., I feel angry), and then to blends of emotions and blends of blends (e.g., I feel slightly jealous, with a dollop of resentment, balanced by hint of admiration). Lane's model suggests emotional awareness follows a developmental trajectory toward greater sophistication across development. But emotion awareness also differs across individuals at any given developmental stage, and critically, it can be improved by training (Subic-Wrana, Bruder, Thomas, Lane, & Köhle, 2005).

Although metacognitive training is an approach that has yet to be applied to the domain of stereotypes and prejudice, it could have important implications. The programs that have been used to improve learning and memory demonstrate that people can sharpen their insight into their mental processes with training, and that cognitive processes that are at one time inscrutable may become easily monitored and more effectively controlled with practice. Although this strategy is one that requires a considerable investment of time and effort, it may hold the greatest promise for overcoming unconscious influences. Whereas most of the solutions we have discussed are enacted on a situation-to-situation basis, the benefits of sharpened meta-cognitive awareness would be experienced across situations. As a result, meta-cognitive training remains an important avenue for future research.

Chapter Summary

We reviewed findings from our own and others' research that call into question the assumption that indirect or implicit measures tap into unconscious thought. Despite the typically low correlation between the attitudes that people explicitly report and those that are revealed on implicit measures, there is relatively little evidence to suggest that this divergence is driven by the existence of attitudes that people harbor subconsciously. Current methodologies fail to provide a critical test of unconscious mental content. Moreover, when factors that are extraneous to the conscious/unconscious distinction are taken into account, the correlation between implicitly and explicitly measured attitudes increases substantially. The motive to conceal attitudes that are viewed as socially unacceptable, the low reliability of many implicit

measures, differences in what people construe as their 'true' attitudes, and poor structural fit between implicit and explicit measures are all factors that can attenuate the correlation. The amplification of the implicit-explicit correlation that occurs when these factors are accounted for suggests that attitudes that are measured implicitly aren't necessarily hidden from consciousness.

While there isn't strong evidence of attitudes and beliefs that linger in the reaches of our unconscious, there is plenty of research indicating that our attitudes and beliefs influence us in countless unconscious ways. Our research has shown that people aren't particularly good at telling when they are versus when they are not being influenced by their biases on a moment-to-moment basis. This disconnect between people's subjective experience of bias and their actual demonstration of it makes it hard to keep one's biases in check. As a result, warning people about the potential for bias or giving them opportunities to refrain from making potentially biased judgments does little to prevent the unwanted influence from occurring.

There may be ways, however, to constrain unconscious influences that deserve further attention. Whereas general consciousness-raising doesn't seem to solve the problem of unconscious influence, strategies that limit the potential for bias and meta-cognitive training may prove to be more effective. By replacing the emphasis on static unconscious attitudes and beliefs with an emphasis on unconscious influences as processes that change over time, debates over unconscious bias can be cast in a new light. There is room for excitement about the fascinating new science of unconscious influence, alongside a more nuanced understanding of what we may and may not be able to achieve, even as cognitive connoisseurs.

References

Akrami, N., & Ekehammar, B. (2005). The association between implicit and explicit prejudice: The moderating role of motivation to control prejudiced reactions. *Scandinavian Journal of Psychology, 46*, 361-366.

Amodio, D., Harmon-Jones, E., Devine, P., Curtin, J., Hartley, S., & Covert, A. (2004). Neural Signals for the Detection of Unintentional Race Bias. *Psychological Science, 15*, 88-93.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or 'Would Jesse Jackson 'fail' the Implicit Association Test?' *Psychological Inquiry, 15,* 257-278.

Baudelaire, C. (1997). *The Parisian prowler: Le spleen de Paris: Petits poemes en prose* (E. K. Kaplan, Trans.). Athens, GA: University of Georgia Press. (Original work published in 1869.)

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*, 674 – 679.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61,* 27-41.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12,* 163-170.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical Versus Actuarial Judgment. *Science, 243,* 1668-1674.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 23,* 316-326.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science, 17*, 383–386.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013-1027.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Reviews in Psychology, 54,* 297–327.

Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, IL: Row, Peterson.

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology, 33*, 573-589.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition, 15,* 485-499.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4-27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369-1385.

Hofmann, W., Gschwendner, T., & Schmitt, M. (2005). On implicit–explicit consistency: The moderating role of individual differences in awareness and adjustment. *European Journal of Personality, 19,* 25–49.

Homer (1999). *The Odyssey* (W. H. D. Rouse, Trans.). New York: Signet Classics.

Jennings, J. M., & Jacoby, L. J. (2003). Improving memory in older adults: Training recollection. *Neuropsychological Rehabilitation, 13,* 417-440.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of*

Personality and Social Psychology, 81, 774-788.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV: Procedures and validity. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies (pp. 59-102)*. New York: Guilford Press.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology, 67,* 808-817.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91-126). New York: Academic Press.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67,* 371-378.

Lane, R. D. (2008). Neural substrates of implicit and explicit emotional processes: A unifying framework for psychosomatic medicine. *Psychosomatic Medicine, 70*, 214-231.

Moskowitz, G.B., Gollwitzer, P.M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77*, 167-184.

Nier, J. A. (2005). How dissociated are implicit and explicit racial attitudes?: A bogus pipeline approach. *Group Processes & Intergroup Relations, 8,* 39–52.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology, 87,* 817-831.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*, 565-584.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81,* 181–192.

Payne, B. K., Burkley, M. A., & Stokes, M. B. (*2008*). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology, 94*, 16-31.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89,* 277–293.

Payne, B. K., Jacoby, L. L., & Lambert, A. J. (2004). Memory monitoring and the control of stereotype distortion. *Journal of Experimental Social Psychology, 40,* 52-64.

Payne, B. K., Lambert, A. J., & Jacoby, L. L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology, 38,* 384–396.

Petty, R. E., Fazio, R. H., & Briñol, P. (in press). Attitudes: Insights from the new wave of implicit measures. Mahwah, NJ: Erlbaum.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75,* 811-832.

Pronin, E., Gilovich, T. D., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review, 111,* 781-799.

Ranganath, K., Smith, C., & Nosek, B. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology, 44*, 386-396.

Rosenthal, R., & Jacobson, L. (1966). Teachers' expectancies: Determinants of pupils' IQ gains. *Psychological Reports, 19,* 115-118.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York: Holt, Rinehart & Winston.

Stewart, B. D., & Payne, B. K. (in press). Bringing automatic stereotyping under control: Implementation intentions as an efficient means of thought control. *Personality and Social Psychology Bulletin*.

Subic-Wrana, C., Bruder, S., Thomas, W., Lane, R., & Köhle, K. (2005). Emotional Awareness Deficits in Inpatients of a Psychosomatic Ward: A Comparison of Two Different Measures of Alexithymia. *Psychosomatic Medicine, 67*, 483-489.

Teong, S. K. (2003). The effect of metacognitive training on mathematical word-problem solving. *Journal of Computer Assisted Learning, 19,* 46-55.

Tonry, M. (1995). *Malign neglect: Race, crime, and punishment in America*. New York: Oxford University Press.

Trawalter, S., & Richeson, J. A. (2006). Regulatory focus and executive function after interracial interactions. *Journal of Experimental Social Psychology, 42,* 406-412.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93*, 239-257.

Trope, Y. & Alfieri, T. (1997). Effortfulness and flexibility of dispositional judgment processes. *Journal of Personality and Social Psychology, 73*, 662-674.

Trope, Y., & Fishbach, A. (2000). Counteractive self-control in overcoming temptation. *Journal of Personality and Social Psychology, 79*, 493-506.

Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science, 16,* 474-480.

Wegner, D. M. (2002). The illusion of conscious will. Cambridge, MA: MIT Press.

Wilson, T. D., & Brekke, N. C. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117-142.

Wilson, T. D., & Nisbett, R. E. (1978). The accuracy of verbal reports about the effects of stimuli on evaluations and behavior. *Social Psychology, 41,* 118-131.

Figure 1.

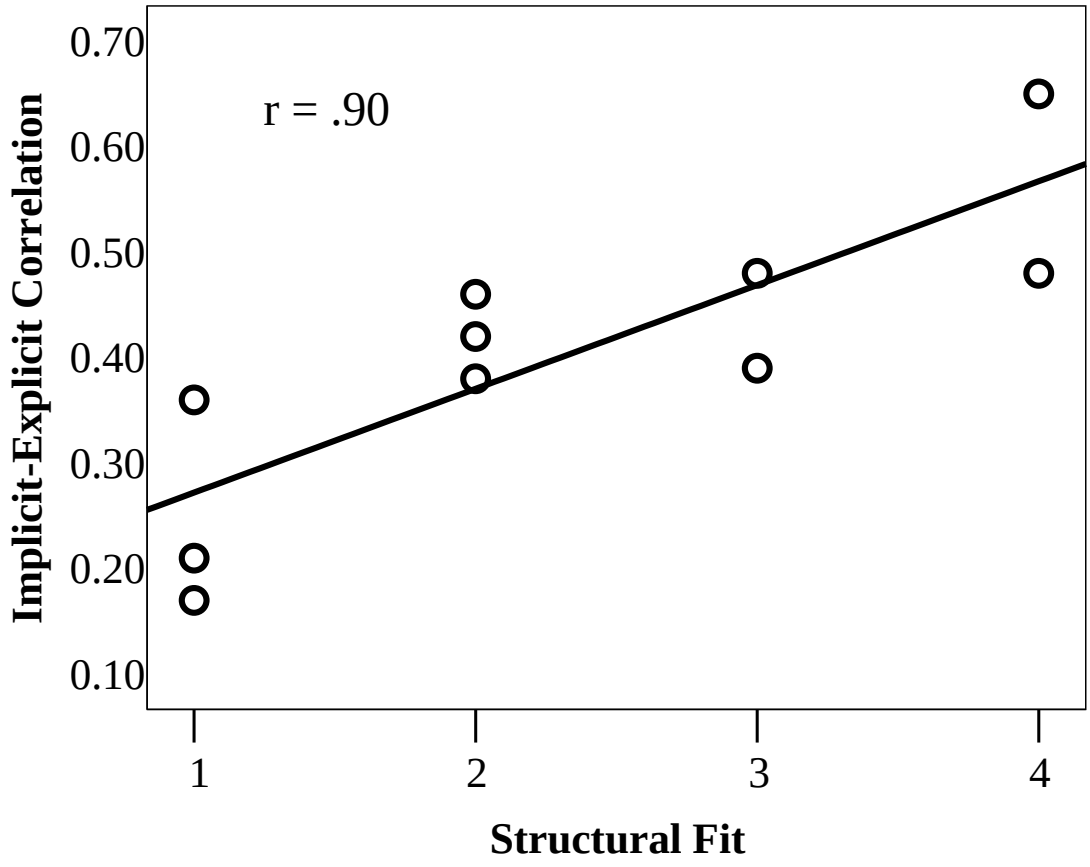
r = .90
Implicit-Explicit Correlation
0.70
0.60
0.50
0.40
0.30
0.20
0.10
1
2
3
4
Structural Fit

Figure 2.

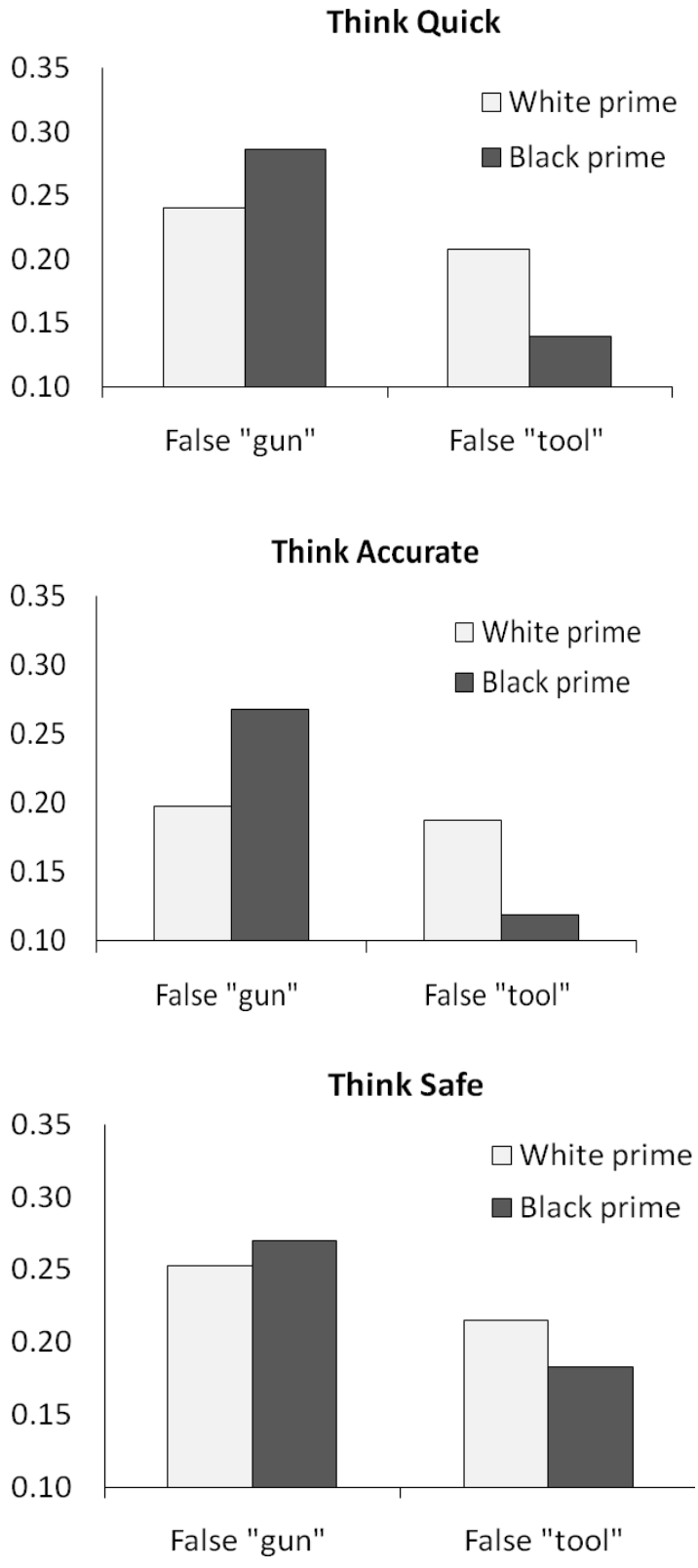
Think Quick
0.35
0.30
0.25
0.20
0.15
0.10
White prime
Black prime
False "gun"
False "tool"
Think Accurate
0.35
0.30
0.25
0.20
0.15
0.10
White prime
Black prime
False "gun"
False "tool"
Think Safe
0.35
0.30
0.25
0.20
0.15
0.10
White prime
Black prime
False "gun"
False "tool"

Figure Caption

Figure 1. The size of the implicit-explicit correlation as a function of structural fit (reverse rank order). Adapted from Payne, Burkley, & Stokes, 2008, Study 2.

Figure 2. False gun and false tool responses in the weapon identification task as a function of three specific intentions. The intention to think “safe” is response to Black faces eliminated the race bias even under fast responding. Adapted from Stewart & Payne (in press).